Evidence of recent signatures of selection during domestication in an Atlantic salmon population

Marine Genomics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Evidence of recent signatures of selection during domestication in an Atlantic salmon population

A.P. Gutierrez [a,b], J.M. Yáñez [c], W.S. Davidson [a,*]

[a] Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada
[b] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, UK
[c] Faculty of Veterinary and Animal Sciences, University of Chile, Av. Santa Rosa, 11735, La Pintana, Santiago, Chile

ARTICLE INFO

Article history:
Received 22 July 2015
Received in revised form 25 November 2015
Accepted 16 December 2015
Available online xxxx

Keywords:
Atlantic salmon
SNP array
Domestication
Signatures of selection
Grilsing

ABSTRACT

Selective breeding practices in Atlantic salmon aquaculture have been carried out intensively since the 1970s. Along with the phenotypic improvement of fish, we expect to observe genomic regions showing evidence of selection for traits related to growth and age at sexual maturation, as well as traits involved in the domestication process. This is mainly linked to the increase in the frequency of favourable alleles at loci that affect the traits of interest in the breeding population. In this study we searched for signatures of selection in the Cermaq Atlantic salmon broodstock, a Mowi strain, which was derived from wild Norwegian populations, and is now farmed in British Columbia, Canada. A 6.5K SNP array was used to genotype 202 fish from the Cermaq population, and the genotypes were compared with four wild populations from Norway. We used three methods based on F_{ST} values to detect signatures of selection. Forty four markers showing divergence in allele frequency were identified as outliers by the three detection methods, suggesting the presence of signatures of selection in the Cermaq population relative to their wild counterparts. Markers identified as outliers are associated with molecular functions that could be related to the selection for economically important traits (e.g., growth) as well as the domestication process (e.g., response to pathogens and environmental stressors). Of particular interest were three outlier markers that had been previously associated with grilsing (i.e., early sexual maturation) an undesirable trait, which has been heavily selected against in Atlantic salmon aquaculture. This study provides clear evidence of the presence of signatures of selection and domestication in a farmed Atlantic salmon population.

1. Introduction

Atlantic salmon genetic improvement has been practiced since the early 1970s when the first breeding trials began in Norway (Gjedrem, 2012). Since then, more than a dozen breeding programs have been established (Rohe et al., 2009), and traits such as growth rate, age at sexual maturity, pathogen resistance, flesh colour and fat content have been included in the breeding goal. The selection responses in salmonids have been higher than in other animal species, due to higher genetic variability and greater fertility, which allow the application of higher selection intensity. For instance, in Norwegian domesticated strains, the response to selection for growth related-traits was found to be greater than 10% in the first few generations (Gjedrem and Baranski, 2010). After more than a half a century of Atlantic salmon aquaculture and even considering the short period of time during which they have experienced intense artificial selection (approximately 12 generations of captive breeding for one Norwegian strain), it can be assumed that the allele frequency of selected loci and the levels of genetic variation have been affected within populations as a consequence of the domestication process. This includes the adaptation to new farming environments and intensive selection for economically important production traits.

Genetic improvement in livestock is mainly driven by increasing the frequency of favourable alleles at loci that affect the traits of interest in populations (Bijma, 2012). The magnitude of these increases is mainly determined by allele substitution effects and allele frequencies at these loci, along with the intensity and accuracy of artificial selection (Falconer and Mackay, 1996). Accordingly, if a population is heavily selected for a particular trait, then there is a higher chance of some alleles reaching fixation. At the same time, aquaculture practices may inadvertently decrease the genetic variation present in farmed stocks. Unless accurate pedigree records are maintained, there is a probability of selecting related individuals as parents, thereby increasing inbreeding.

A series of events may occur in genomic regions of populations affected by intensive selection. For instance a “hard sweep” is the process that occurs when the genetic variation in selected regions is disrupted leading to an association between an adjacent locus and the selected site. On the other hand, there is another scenario known as a “soft sweep”, in which more than one positive allele can be present within

* Corresponding author.
E-mail address: wdavidso@sfu.ca (W.S. Davidson).

http://dx.doi.org/10.1016/j.margen.2015.12.007

the selected locus and a drastic reduction in genetic variation in the genomic region does not occur. Soft sweeps involve variants at a selected locus, and therefore they may produce many different haplotypes at the closely linked sites. Additionally, when adaptation occurs by selection of polygenic traits, it generally induces an increase in the allelic frequency of several loci, which have a favourable effect on the trait. These alleles do not necessarily reach fixation and the haplotype pattern corresponds to several partial selection signatures or multiple "partial sweeps" (Pritchard et al., 2010). Signatures of selection are genomic regions which contain DNA sequences affecting genetic variation of characters that have undergone natural or artificial selection (Qanbari et al., 2012). Such signatures may be identified using genomic information and different analytical approaches (Lopez et al., 2015). This is possible because selection can affect the DNA sequence at a particular region. This is known as the "hitchhiking effect" (Smith and Haigh, 1974). This leaves a "footprint" or "signature" around the selected gene variant, which yields a specific and detectable genomic sequence pattern (Pennings and Hermisson, 2006).

Previous studies have shown evidence of directional selection in genomic regions of farmed Atlantic salmon when compared to their wild counterparts (Vasemägi et al., 2012) or when comparisons are made among pairs of farmed populations (Martinez et al., 2013), even when a limited number of markers is used. Using a SNP chip, Karlsson et al. (2011) were able to identify a panel of 60 markers capable of differentiating farmed Atlantic salmon from wild populations in Norway. However, they did not provide evidence of significant genetic differentiation between them. Based on these observations it has been suggested that genetic differences on a genomic scale, between Norwegian farmed and wild populations are the result of small allele frequency changes at a large number of loci, rather than large allele changes at few loci.

The development of technologies for typing dense marker genotypes provides the opportunity to simultaneously analyse thousands of SNPs and more precisely identify regions of the genome that show evidence of selection. These tools have already been used for the analysis of many livestock species such as cattle, sheep, pig and chicken (Ai et al., 2013; Kijas et al., 2012; Qanbari et al., 2012, 2014; Ramey et al., 2013), by performing genome-wide scans comprising from hundreds of thousands to millions of SNPs. The identification of selection signatures may help to unravel the genetic factors and mechanisms involved in important biological traits, because these regions might have adaptive and functional relevance underlying their selection (Nielsen et al., 2007). Atlantic salmon provide an excellent model for studying the effects of early selection and the domestication process, as some of their populations have been domesticated very recently in their evolutionary history and there is the availability of both domesticated and wild populations from the same lineage. This study was designed to detect signatures of domestication and early selection processes in Atlantic salmon using an Illumina iSelect SNP-array (Kent et al., 2009) in order to get a better understanding of Atlantic salmon adaptation to a farming environment and high selection pressures for production traits.

2. Materials and methods

2.1. Farmed fish samples and wild fish genotype data

The Atlantic salmon used in these analyses were part of a commercial broodstock program initiated by Cermaq Canada in 1995 and based on the Mowi strain of Atlantic salmon (imported to British Columbia, Canada in the mid 1980s). The Mowi strain was established in the late 1960s with a major contribution from the River Bolstad, in the Vosso watercourse, and the River Aaroy with additional contributions from wild salmon captured in the sea near Osterfjord and Sotra in western Norway (Glover et al., 2009; Verspoor et al., 2007). Considering the time since the Mowi strain was established, we estimate that it has gone through approximately 12 generations of selection, seven of these in Norway and five in Canada. The Cermaq population has been selectively bred for faster growth and reduced early sexual maturation. We chose 202 parents from the 2005 broodstock year. DNA from all 202 fish was obtained from fin-clips using the DNeasy Blood & Tissue Kit (Qiagen) and then sent for genotyping. SNP genotyping was carried out at CIGENE, Norwegian University of Life Sciences, Ås using an Atlantic salmon 6.5K Illumina iSelect SNP-array (Kent et al., 2009), described by Lien et al. (2011). The level of relatedness of the 202 fish was estimated from a genomic kinship matrix based on identity-by-state (IBS) values obtained using the *GenABEL* package implemented in R. The mean value was −0.0025 with a median of −0.0069, indicating low levels of relatedness.

To perform comparative analyses we used the genotype data from four Norwegian wild Atlantic salmon populations available at the Dryad Digital Repository (http://datadryad.org), obtained and made publicly available by Bourret et al. (2013). The wild populations were identified according to the region of origin (Fig. 1), as follows: Tana (TAN) with 29 individuals (1), Gaula (GAU) with 43 individuals (2), Laerdaselva (LAR) with 25 individual (3) and Numedalslagen (NUM) with 43 individuals (4), giving a total of 140 wild samples, which together comprised the WILD dataset. More details regarding these wild populations can be found in Bourret et al. (2013). These populations were chosen to provide a good representation of the wild populations in Norway (based on their geographical provenance) for comparisons with the Mowi strain, which can be considered as initially being a composite strain established with contributions from different Norwegian rivers.

For comparative purposes, the data-sets from both, the WILD set and the Cermaq populations were filtered to keep only the shared markers. These markers were used in the subsequent analyses, after filtering for a call rate threshold of 95%.

2.2. Basic population genetic statistics and structure

Deviations from Hardy–Weinberg equilibrium (HWE) were tested with the exact test (Guo and Thompson, 1992), as implemented in GENEPOP 3.4 (Rousset, 2008) and Arlequin 3.5 (Excoffier and Lischer, 2010). Genetic differentiation between populations was measured with pairwise F_{ST} estimates (Weir and Cockerham, 1984), using Arlequin 3.5 (Excoffier and Lischer, 2010). Genetic distances between populations were estimated based on Nei (1972), implemented in the R package *adegenet* (Jombart, 2008). Inbreeding coefficients (F_{IS}) were calculated using the R package *Demerelate* (Kraemer and Gerlach, 2013).

Population structure was inferred from the SNP markers using a hierarchical Bayesian modelling construct in the program *STRUCTURE* (Pritchard et al., 2000; Falush et al., 2007), using a burn-in of 100,000 iterations and running 100,000 iterations ($K = 2$). Then we conducted an individual-based principal component analysis (PCA) implemented in the R package *adegenet* (Jombart, 2008). We used the function *find.clusters* to estimate the optimal number of groups with the Bayesian Information Criterion (BIC) method, and we used the function *a.score* to determine the optimal number of discriminant functions to retain (Jombart et al., 2010).

2.3. Detection of F_{ST} outlier loci

Three different tests for the detection of loci subject to directional selection during domestication were used. These tests are based on different assumptions, but rely on the rationale that directional selection increases genetic differentiation between populations and reduces variation at linked loci, providing additional support for the identification of outlier loci. For the first stage of the analyses, the data from the four wild Atlantic salmon populations were grouped into one set of 140 samples (WILD) that was compared against the dataset from the 202 farmed samples.

Please cite this article as: Gutierrez, A.P., et al., Evidence of recent signatures of selection during domestication in an Atlantic salmon population, Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.12.007

Fig. 1. Map showing sample sites of Norwegian wild populations. Numbers 1, 2, 3 & 4 represent Tana (TAN), Gaula (GAU), Laerdaselva (LAR) and Numedalslagen (NUM) populations, respectively.

The first method used was implemented in Arlequin 3.5 (Excoffier and Lischer, 2010). This program uses a similar algorithm to the n-island model of FDIST (Beaumont and Nichols, 1996) but accounts for historical meta-population structure with a hierarchical island model (H) (Excoffier et al., 2009), which aims to reduce the number of false positive F_{ST} outlier loci. Arlequin runs based on the assumption that the average migration rate between populations on different islands is lower than that between demes on the same island and also that the heterozygosity between populations can be inferred using the heterozygosity within a population (Excoffier and Lischer, 2010). Significance of outlier SNP markers was assessed by running 50,000 simulations, 100 demes and 20 groups.

The second method used to detect footprints of selection also uses a hierarchical-Bayesian model that is similar to the original F_{ST}-outlier test, FDIST (Beaumont and Nichols, 1996). Using a logistic regression model, F_{ST} is broken down into two effects: a locus effect and a population effect, relaxing the assumption of a symmetrical island model by allowing for population structure asymmetries and implemented in BayeScan (Foll and Gaggiotti, 2008). The program calculates the Posterior Odds (PO), from the posterior probability of a particular locus being under selection using a value of prior odds equivalent to 10:1 by comparing models with and without a selection term and by using the proportion of loci with a strong increase in F_{ST} relative to other loci among the Markov Chain Monte Carlo outputs of its simulations. The analyses were run with the default settings and a locus was considered to be under strong selection if log10 (PO) was ≥2 and under decisive selection if log10 (PO) was ≥3.

The third method we used was to run the analysis using Arlequin 3.5 (Excoffier and Lischer, 2010), but not using the hierarchical island model (nH), which would only simulate a finite island model. The analysis simulations were run using the same specifications used for the Arlequin (H) method.

For a second stage, we performed F_{ST} outlier analyses for the Cermaq population compared to each of the four wild populations (Fig. 1). In this case, only BayeScan was used for analyses, using the same setting as in the previous analysis.

2.4. SNP annotation and Gene Ontology

Markers were assigned to their respective chromosomal position according to the previously described Atlantic salmon linkage map, which contains ~5650 SNPs and was constructed using genotyping data from 143 families comprising 3297 fish (Lien et al., 2011). This map contains 29 linkage groups, each of which was assigned to its specific chromosome according to the nomenclature established by Phillips et al. (2009). Markers that were not located on the map were only assigned

Please cite this article as: Gutierrez, A.P., et al., Evidence of recent signatures of selection during domestication in an Atlantic salmon population, Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.12.007

to a chromosome by identifying the Atlantic salmon genome contig containing them, available at www.asalbase.org.

The nucleotide sequences corresponding to the SNPs that showed a significant association were compared by BLAST against information from the Atlantic salmon genome sequencing project v1.0 (Davidson et al., 2010), which is publicly available at ASalBase (www.asalbase.org) and NCBI (http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AGKD). SNP markers were then assigned to a specific whole genome shotgun (WGS) contig by sequence similarity searches. WGS contigs were annotated using an in-house annotation pipeline (trutta.mbb.sfu.ca) and BLASTn alignments.

Blast2go (Conesa et al., 2005) was used to obtain the Gene Ontology (GO) annotation (Ashburner et al., 2000). Homology searching was first realized through a BLAST search of the available flanking sequences for each SNP on the NCBI nr public database with the e-value threshold set to $1 \times 10E-10$. Blast2go then retrieved GO terms associated with the BLAST hits.

3. Results

3.1. Data management

Genotype data from wild and farmed samples were compared and organized for comparative purposes. Data from wild populations obtained from Bourret et al. (2013) contained genotypes from 6176 markers, whereas the Cermaq dataset only contains 5568 markers. After comparison, a total of 5088 markers were shared between both datasets. Filtering for a minimum call rate (>0.95), showed that 4905 markers passed the threshold among the datasets.

3.2. Population differentiation and genetic structure

Within population observed and expected heterozygosity levels were similar across the four wild and the farmed populations (Table 1). The dataset comprising all wild populations (WILD) however, showed a slightly lower level of observed heterozygosity (0.331) than expected (0.338). The heterozygosity levels of the Cermaq (farmed) population were lower than the individual wild populations but higher than the WILD population dataset. On the other hand, inbreeding values (F_{IS}) show that all populations show low or negative values of inbreeding (Table 1).

Between population comparisons based on F_{ST} values show low levels of differentiation between the wild populations. It should be noted however, that the LAR, NUM and GAU populations showed a lower level of differentiation between one another (F_{ST} 0.017–0.040) than they did to the TAN population (F_{ST} 0.048–0.065) (Table 2). This observation can be explained by considering the geographical locations of these populations (Fig. 1). LAR, NUM and GAU are all Atlantic Ocean populations whereas TAN is a Barents-White Sea population. Comparison of the LAR, NUM and GAU populations against the Cermaq population showed a lower level of differentiation (F_{ST} 0.072–0.083) than did the TAN population (F_{ST} 0.116). Additionally, Nei genetic distance values showed a similar pattern in terms of the distance between TAN and the other wild populations, and also their distance to the Cermaq population. The Cermaq population appears to be genetically more distinct compared to all four Norwegian wild populations (Table 2 and Fig. S1).

Structure analysis showed the presence of two main clusters differentiating the four wild populations from the Cermaq population, but not from one another (Fig. 2). Similarly, PCA analysis revealed three clusters, according to the lowest BIC, classified along the two first principal components, which explained 64.13% and 25.97%, respectively of the total genetic variation among individuals (Fig. 3).

Table 1
Genetic diversity (heterozygosity), Inbreeding coefficient (F_{IS}) and Hardy Weinberg Equilibrium (HWE) values for the analysed populations.

Population	Genetic diversity		F_{IS} (p-value)	HWE (mean p-value)
	H_o	H_e		
TAN	0.351(0.176)	0.344(0.149)	0.036(0.278)	0.508
GAU	0.351(0.159)	0.352(0.145)	0.0021(0.278)	0.517
LAR	0.358(0.165)	0.357(0.145)	−0.0037(0.303)	0.513
NUM	0.354(0.171)	0.346(0.148)	0.041(0.266)	0.499
Cermaq	0.346(0.160)	0.340(0.151)	−0.008(0.247)	0.481
WILD	0.332(0.157)	0.339(0.153)	0.049(0.218)	0.457

Observed (H_o) and expected (H_e) heterozygosity values (with standard error in parentheses). F_{IS} (Inbreeding coefficient, mean p-values are shown in brackets indicating low level of significance. HWE mean p-values for each population.

Table 2
Population differentiation values based on Nei genetic distances and F_{ST}.

Population	Population differentiation					
	TAN	GAU	LAR	NUM	Cermaq	WILD
TAN	0	0.0348	0.0451	0.0409	0.0699	–
GAU	0.0482	0	0.0171	0.0224	0.0401	–
LAR	0.0644	0.0178	0	0.0313	0.0413	–
NUM	0.0660	0.0274	0.0404	0	0.0512	–
Cermaq	0.1161	0.0722	0.0728	0.0833	0	0.039
WILD	–	–	–	–	0.0692	0

Nei genetic distance values are shown above the diagonal. F_{ST} values are shown below the diagonal. All F_{ST} values were significant ($p < 0.001$).

3.3. Outlier marker detection between the WILD dataset and Cermaq population

Outlier detection analyses were carried out using two groups: the first comprised the genotypes from the four wild populations (LAR, NUM, TAN and GAU, 140 individuals referred to as WILD), and the second the genotypes from the Cermaq population (202 individuals). Analyses were performed using two different programs associated with three approaches. From these three approaches, 44 markers were identified as outliers by all three methods (Table 3 and Fig. 4). In particular, analysis performed using BayeScan detected 46 outlier markers. Of these, 31 showed evidence of decisive selection (log10 (PO) ≥ 3) and four showed strong evidence of selection (log10 (PO) = 2–2.99). The remaining markers showed little evidence of selection even though they were detected as outliers (Table 3). The second approach, using Arlequin (H) detected 105 markers showing apparent evidence of selection (Table S1), and only five markers were uniquely detected as outliers using this method. The third approach, using Arlequin (nH), detected 107 markers as outliers according to their p-value ($p < 0.01$). This method detected the highest number of outliers, as shown in Table S2 and Fig. 4. Of these, 98 were common to the Arlequin H (Hierarchical model) method, and only nine markers were not detected by any of the other methods (Fig. 4). A graphical representation of the outlier detection obtained from Arlequin and BayeScan methods is shown in Fig. S2.

3.4. Outlier marker detection between the four wild populations and the Cermaq population

Given the robust conservative results that BayeScan gave for the analysis of the WILD dataset, we decided to use it for the detection of outliers in the Cermaq population compared to each of the four wild populations from Norway. The analyses provided similar results to the analysis performed using the dataset of the combined wild populations. Numerous markers detected as outliers for the WILD dataset were also detected in this analysis (Fig. 5 and Table S3). Of the 44 markers detected as outliers by the three methods used in the previous analysis using the WILD dataset (Fig. 4), 30 were also detected when comparing the Cermaq population with the Tana (TAN) population, 40 with the Gaula (GAU) population, 35 with the Laerdaselva (LAR) population

Please cite this article as: Gutierrez, A.P., et al., Evidence of recent signatures of selection during domestication in an Atlantic salmon population, Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.12.007

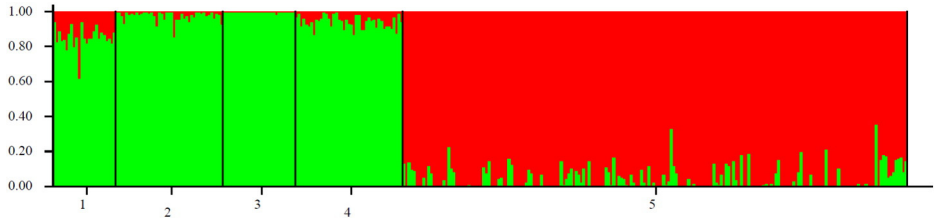

Fig. 2. Analysis of the structure of the five populations (1 TAN, 2 GAU, 3 LAR, 4 NUM, (1–4 = WILD), 5 Cermaq (farmed)), according to STRUCTURE (K = 2).

and 39 with the Numedalslagen (NUM) population. The complete list of markers identified as outliers analysing each of the four populations can be found in Tables S4–S7 and the graphical representation of the BayeScan detection is shown in Fig. S3.

We found that there were 26 common markers detected as outliers when the analyses were performed on each of the four wild populations (Fig. 5). These 26 markers were also detected as outliers when using the WILD dataset (and are part of the 44 markers identified as outliers in Fig. 4). Comparison with the LAR population detected 37 outliers, two of which were only detected for this population. For the NUM population, 44 markers were identified as outliers with five markers only detected in this population. For the GAU population, 45 markers were detected as outliers and six of them were specific for this population. The TAN population analysis detected 45 markers as outliers, but of these, fifteen were specific to this population. For the GAU population 33 markers showed at least strong evidence of selection (log (PO) ≥ 2). There were 29 such markers for the LAR population, 30 such markers for the NUM population and 20 such markers for the TAN population. Of these markers, 16 were shared between the four populations. As expected, all of these 16 markers also showed strong evidence of selection when the WILD dataset was analysed.

3.5. Genomic regions under selection and their putative associated genes

Atlantic salmon chromosome 11 (Ssa11) contains the highest number of markers showing at least strong evidence of selection (five), followed by Ssa02, SSa04 and Ssa06 each with three markers (Table 4). Markers located on Ssa11, however, span a wide region along the chromosome (according to the female map; Lien et al., 2011), a situation observed for all chromosomes.

BLASTn annotation of the sequences containing the SNPs (or the surrounding sequence based on the WGS contig) gave significant hits for 30 of the 44 markers that showed evidence of selection (Table 4). GO annotation of these markers on the other hand gave a lower number of results, only assigning GO terms for 21 markers (Table S8). GO terms segregated into many levels of biological processes, molecular functions and cellular component.

4. Discussion

In the present study, ~5000 SNP markers were compared in four wild populations from Norway and one farmed population from British Columbia, Canada, originating from a Norwegian strain. Several genomic regions and 44 SNP markers potentially affected by artificial selection and associated with domestication in Atlantic salmon were detected by three different methods. Contrary to previous studies seeking to identify signatures of selection (Karlsson et al., 2011; Mäkinen et al., 2015; Vasemägi et al., 2012), these results suggest that the footprints of selection in Atlantic salmon (at least those farmed in the west coast of Canada) are significant after only ~12 generations of independent domestication and high selection pressure for economically important traits (e.g., rapid growth and non-grilsing).

4.1. Population structure and genetic differentiation

Analysis of the genetic structure of the populations showed that the levels of heterozygosity of the Cermaq population were similar to those of wild populations (see Table 1) despite the possible effects of artificial selection, potential inbreeding and the loss of alleles due to genetic drift. The small excess of heterozygosity found in most populations is not significant with respect to HWE values. Estimates of inbreeding (F_{IS}), like

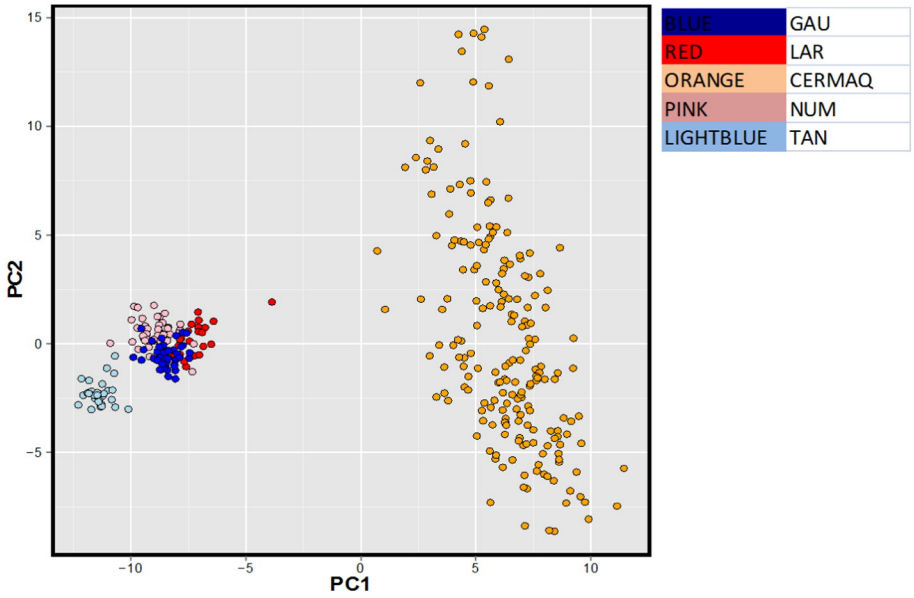

Fig. 3. Principal component analysis (PCA) representing the structure of the analysed populations.

Table 3
Summary of markers identified as outliers and shared by the three utilized methods.

Rank	SNP ID	BayeScan			Arlequin (nH)		Arlequin (H)	
		Log10 (PO)	alpha	F_{ST}	Obs F_{ST}	p-Value	Obs F_{ST}	p-Value
1	ESTNV_35725_1965	∞	3.0312	0.4951	0.9852	1.00E−07	0.9852	1.00E−07
1	ESTNV_34596_367	∞	2.971	0.4837	0.976	1.00E−07	0.976	1.00E−07
1	ESTV_12965_608	∞	2.8258	0.4563	0.9402	1.00E−07	0.9402	1.00E−07
1	ESTNV_14050_192	∞	2.4394	0.3843	0.7516	1.00E−07	0.7516	1.00E−07
1	GCR_cBin17565_Ctg1_349	∞	2.3681	0.3716	0.7151	1.00E−07	0.7151	1.00E−07
1	GCR_cBin21476_Ctg1_139	∞	2.3157	0.362	0.7019	1.00E−07	0.7019	1.00E−07
1	ESTNV_35581_518	∞	2.2856	0.3573	0.8731	1.00E−07	0.8731	1.00E−07
1	GCR_cBin49985_Ctg1_148	∞	2.2572	0.3514	0.928	1.00E−07	0.928	1.00E−07
1	ESTNV_27648_141	∞	2.2298	0.3476	0.8442	1.00E−07	0.8442	1.00E−07
1	ESTNV_23463_496	∞	2.1981	0.3411	0.9164	1.00E−07	0.9164	1.00E−07
1	ESTNV_33653_51	∞	2.1699	0.3367	0.8694	1.00E−07	0.8694	1.00E−07
1	ESTV_17052_125	∞	2.1193	0.328	0.866	1.00E−07	0.866	1.00E−07
1	ESTNV_28438_568	∞	2.1191	0.3278	0.8604	1.00E−07	0.8604	1.00E−07
1	GCR_cBin6804_Ctg1_99	∞	2.0867	0.3216	0.8934	1.00E−07	0.8934	1.00E−07
1	ESTNV_35410_1102	∞	2.0355	0.3135	0.5381	1.00E−07	0.5381	1.00E−07
1	ESTNV_31412_354	∞	2.0211	0.3124	0.7356	1.00E−07	0.7356	1.00E−07
1	GCR_cBin28815_Ctg1_209	∞	1.9957	0.3071	0.5425	1.00E−07	0.5425	1.00E−07
1	ESTNV_24005_495	∞	1.9884	0.3045	0.8611	1.00E−07	0.8611	1.00E−07
1	GCR_cBin5494_Ctg1_253	∞	1.979	0.304	0.5041	1.00E−07	0.5041	0.0003
1	ESTV_15617_513	∞	1.9646	0.3019	0.4937	1.00E−07	0.4937	0.0011
1	GCR_cBin6555_Ctg1_258	∞	1.9617	0.3013	0.4904	1.00E−07	0.4904	0.0014
1	GCR_cBin47868_Ctg1_188	∞	1.9323	0.2965	0.4754	1.00E−07	0.4754	0.0021
1	ESTNV_12926_381	∞	1.9058	0.2915	0.8187	1.00E−07	0.8187	1.00E−07
1	ESTV_19580_277	∞	1.8758	0.2874	0.4443	1.00E−07	0.4443	0.0011
1	ESTNV_36261_377	∞	1.8809	0.2871	0.8328	1.00E−07	0.8328	1.00E−07
1	ESTNV_30332_660	∞	1.8746	0.2857	0.8163	1.00E−07	0.8163	1.00E−07
1	ESTV_19974_640	∞	1.8162	0.2762	0.7837	1.00E−07	0.7837	1.00E−07
1	BASS121_B7_G07_685	∞	1.7838	0.2715	0.7896	1.00E−07	0.7896	1.00E−07
1	GCR_cBin27948_Ctg1_211	∞	1.7463	0.2651	0.7263	1.00E−07	0.7263	1.00E−07
2	GCR_cBin3425_Ctg1_442	3.699	1.7197	0.2611	0.7351	1.00E−07	0.7351	1.00E−07
3	ESTNV_28566_397	3	1.7655	0.2699	0.4057	1.00E−07	0.4057	0.00107
4	GCR_cBin31103_Ctg1_204	2.584	1.7522	0.2687	0.3821	1.00E−07	0.3821	0.00036
5	ESTNV_30600_327	2.375	1.5625	0.2385	0.6772	1.00E−07	0.6772	1.00E−07
6	GCR_cBin108_Ctg1_303	2.335	1.6731	0.2562	0.3443	1.00E−07	0.3443	1.00E−07
7	ESTNV_14967_182	2.317	1.6844	0.258	0.3564	1.00E−07	0.3564	5.95E−06
8	ESTV_16140_475	1.628	1.5616	0.2402	0.3161	0.000201	0.3161	1.00E−07
9	ESTNV_35345_1492	1.243	1.4723	0.229	0.2788	0.003297	0.2788	0.00037
10	GCR_rBin19088_Ctg1_157	1.047	1.4771	0.2326	0.4323	0.000972	0.4323	0.00044
11	GCR_cBin33290_Ctg1_231	0.813	1.1395	0.1835	0.5136	0.000226	0.5136	1.00E−07
12	ESTNV_36128_1360	0.624	1.1232	0.1849	0.5111	0.00102	0.5111	4.83E−06
13	GCR_cBin31103_Ctg1_155	0.59	1.0615	0.1758	0.3372	1.00E−07	0.3372	0.00029
14	GCR_cBin1536_Ctg1_164	0.535	1.0192	0.1714	0.5186	0.000357	0.5186	0.00069
15	ESTV_20616_997	0.431	1.106	0.1882	0.5186	0.000357	0.3728	0.00159
16	GCR_cBin11246_Ctg1_207	0.315	0.9259	0.1641	0.5032	0.000973	0.5032	9.95E−05

Rank order is based on BayeScan significance levels.

the heterozygosity levels showed a higher value on the WILD dataset than in the Cermaq population and in any of the wild populations, although this was not significant (Table 1). Population differentiation based on F_{ST} and Nei distances showed similar values between the populations. TAN was an out-group to the other three Norwegian wild populations, and the Cermaq population was genetically more distant with respect to all of the Norwegian wild populations (Table 2 and Fig. S1).

PCA and structure analyses indicate that the Cermaq population has become genetically distinct from the natural populations in Norway, possibly due to the effect of selective pressures. In particular the PCA analysis (Fig. 3), which was constructed from the two largest dimensions of variation (PC1 and PC2), explained 64% and 26% of the total variance, respectively. GAU, LAR, NUM and TAN, display a low variation for PC1 and PC2 when compared to the Cermaq population, especially when we observe the variation of this population for PC2. The close clustering of wild Norwegian populations and the divergence from the Cermaq population makes sense from the demographic perspective. The highest variation of the farmed population for PC1 and, especially, for PC2, when compared to the wild populations can be attributed to an admixture process when the base population of Mowi was created. The breeding between individuals from different populations most certainly generated a higher level of genetic variation for this composite population when compared to their ancestral progenitor populations. Our analysis demonstrates that much of this original genetic variation is still present in the Cermaq population, which is the result of a deliberate breeding strategy designed to maintain the levels of genetic variation and genetic diversity.

4.2. Detection of selection signatures

Three methods were used for the detection of outliers showing evidence of selection in the Cermaq population. In general, when using the combined WILD dataset, the three simulation-based methods gave similar results agreeing on the identification of 44 markers as outliers. Arlequin (nH) gave the highest number of markers identified as outliers (107), followed by Arlequin (H) analysis with (105), while BayeScan gave the lowest (46). This difference in the number of markers could be either due to the inability of the BayeScan methods to detect real signatures of selection at these loci or simply because a large number of the outlier loci detected by Arlequin 3.5 are false positives (type I error); the latter being more likely (Narum and Hess, 2011).

The analysis of outliers based on individual populations also identified numerous significant markers, of which the majority were shared among the different comparisons. The number of outliers identified in

Please cite this article as: Gutierrez, A.P., et al., Evidence of recent signatures of selection during domestication in an Atlantic salmon population, Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.12.007

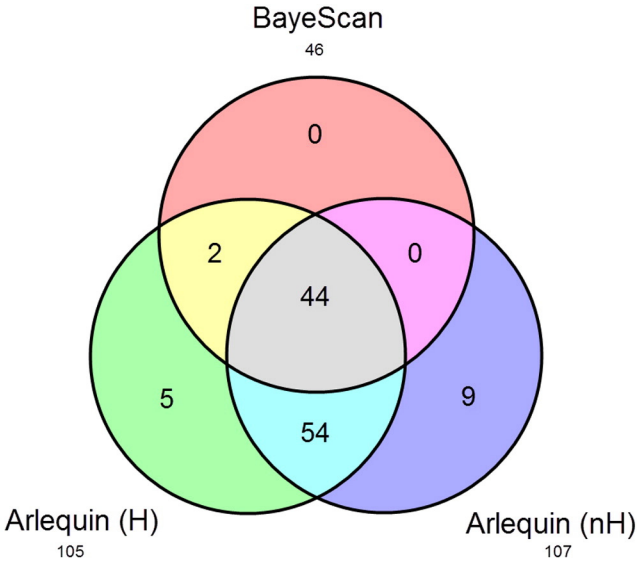

Fig. 4. Venn diagram illustrating the similarities and differences in the detection of outliers when the Cermaq farm population was compared with a dataset comprising the four wild Norwegian populations (WILD) using two approaches implemented in Arlequin (H: hierarchical island model; nH: finite island model) and the BayeScan method.

the comparison of the Cermaq farmed Atlantic salmon against the GAU, NUM and LAR populations was similar, ranging from 37 to 45 markers, and most of them were also detected in the previous analysis using the combined WILD dataset (Table S3). In contrast, when the analysis was based on the TAN population 45 outliers were detected, but only 31 of these were found in the WILD dataset comparison. This situation was not totally unexpected given the different geographical origin of these populations and the known structure of the populations (Figs. 1, 2 and 3). While LAR, GAU and NUM populations can be classified as Atlantic Ocean populations, the TAN population represents a Barents-White Sea population, geographically distant from the others. This was also observed in the population differentiation values shown in Table 1, where the TAN population shows the highest differentiation values when compared to the other populations and with the farmed (Cermaq) population. These results may reflect differences between the four wild populations as a result of divergent local adaptation, thus leading to different outlier markers when comparing each one of them against the

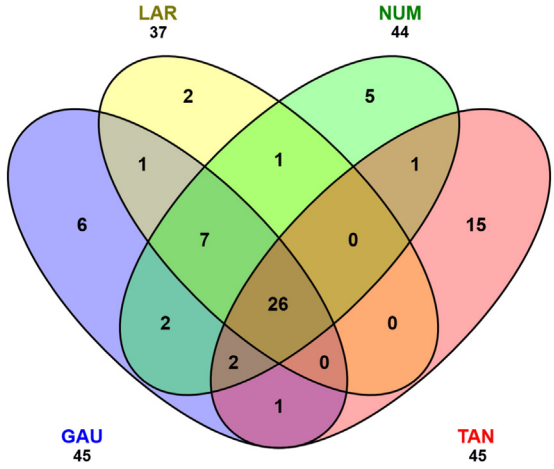

Fig. 5. Venn diagram showing the similarities and differences in outlier detection between the four wild Norwegian populations and the farmed population. In this case the Cermaq population was analysed against each of the four wild Norwegian populations independently using the methods implemented in BayeScan software.

Cermaq population and with one another. Nevertheless, 56% of the outlier markers (26 out of 46) were consistent between the two analyses, Cermaq against WILD and Cermaq against each wild population separately, using BayeScan.

Results of outlier tests for identifying loci under possible selection must be evaluated with care since simulation studies have identified considerable variation in false-positive and false-negative rates among different F_{ST} outlier methods, with BayeScan having the lowest false-positive rate (Excoffier et al., 2009; Narum and Hess, 2011). Some authors have suggested that only those outliers supported by multiple statistics should be considered as the most promising candidates. Additionally, numerous drawbacks have been identified while analysing for signatures of selection. For example, deviations from the demographic models assumed in the tests are usually found (Excoffier et al., 2009). There could also be an inclusion of severely bottlenecked populations, which would lead to high numbers of false positives (Teshima et al., 2006). Another factor could be differences in mutation rates among loci (Beaumont and Nichols, 1996). It has also been proposed that the strong correlation in co-ancestry found in rivers, oceanic ridges or coastlines increases the neutral variance in F_{ST}, resulting in a high rate of outliers (Fourcade et al., 2013), which could be the case of our analysed populations. We chose to use three different detection approaches to overcome these potential problems.

Previous studies analysing selection in Atlantic salmon strains have been based on a limited number of markers (e.g., microsatellites and EST-derived markers), and these have detected few regions potentially under selection, supported by multiple outlier tests (Vasemägi et al., 2005, 2012). Variable situations have been observed in many other species (reviewed by Nosil et al., 2009). The use of high density SNP arrays has become more popular, and along with the increase of marker density and therefore the linear coverage of the genome, a higher probability of detecting genomic events as signatures of selection is expected. Its use has already made possible the identification of signatures of natural selection in Atlantic salmon populations (Bourret et al., 2013; Perrier et al., 2013; Zueva et al., 2014); however, the analysis of domestication seems to be different. Recently, Mäkinen et al. (2015) analysed three populations of Atlantic salmon, one from North America (Saint John, Canada) and two from Europe (Ireland and Sweden) using the 6.5K SNP array used in this study, and found little evidence for signatures of domestication using the BayeScan and standardized heterozygosity (Kauer et al., 2003) methods. The results presented here indicate a different scenario, with numerous outliers suggesting signatures of selection occur on most of the Atlantic salmon chromosomes (Table 4). In terms of the number of markers identified as outliers, approximately 1% of the total of markers analysed indicated evidence of selection. Differences in the outcome of the analyses may be due to various reasons: First, the populations analysed by Mäkinen et al. (2015) (same populations used by Vasemägi et al., 2012) are from different geographical locations, which could represent differences in the levels of genetic variation and standing variation found in the Cermaq population. Second, only the Saint John population represents a population under selection for aquaculture purposes. The other two (from Ireland and Sweden) are stock supplementation populations that do not go through strong artificial selection. In contrast, the Cermaq population analysed in this study has been heavily selected for economically important production traits over ~12 generations. Third, the Saint John population belongs to the North American salmon lineage, and it has been shown that North American and European Atlantic salmon are genetically distinct, with differences in chromosome numbers and chromosomal rearrangements (Lubieniecki et al., 2010; Brenna-Hansen et al., 2012). In this study, the use of a few thousand markers distributed across all of the 29 Atlantic salmon chromosomes provides enough resolution to detect loci under selection. Still, it is important to note that differences found between the natural populations from Norway and the Cermaq population could carry a bias associated with independent evolutionary trajectories where both drift and selection may have occurred independently.

Please cite this article as: Gutierrez, A.P., et al., Evidence of recent signatures of selection during domestication in an Atlantic salmon population, Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.12.007

Table 4
Linkage map position and annotation of the markers showing strong to decisive evidence of selection in the within population analysis.

Rank	SNP ID	Chromosome	Female map	Male map	WGS contig	Annotation
1	ESTNV_31412_354	Not assigned	–	–	AGKD01092464	PREDICTED: *Esox lucius* beta-1,4-galactosyltransferase 1-like
1	ESTNV_35581_518	Ssa01	–	–	AGKD01044862	*Salmo salar* AN1-type zinc finger protein 5
1	ESTNV_30332_660	Ssa01	52.7	48.3	AGKD01224489	No hits
1	GCR_cBin6804_Ctg1_99	Ssa02	86.2	24.9	AGKD01100126	No hits
1	ESTNV_35410_1102	Ssa02	70.6	23.8	AGKD01075420	*Salmo salar* FXYD domain containing ion transport regulator 5a (fxyd5a)[a]
1	ESTNV_24005_495	Ssa02	–	–	AGKD01164226	PREDICTED: *Esox lucius* nucleotide-binding oligomerization domain containing 1 (nod1)
1	ESTNV_33653_51	Ssa03	–	–	AGKD01200451	PREDICTED: *Larimichthys crocea* actin-related protein 2/3 complex subunit 1A-like
1	ESTNV_28438_568	Ssa03	–	–	AGKD01023149	*Salmo salar* NAD-dependent deacetylase sirtuin-5
1	ESTNV_34596_367	Ssa04	–	–	AGKD01003060	PREDICTED: *Esox lucius* junctional adhesion molecule C-like
1	GCR_cBin17565_Ctg1_349	Ssa04	110	67.4	AGKD01411305	No hits
1	GCR_cBin6555_Ctg1_258	Ssa04	20.5	2.1	AGKD01238521	No hits
1	ESTNV_35725_1965	Ssa05	–	–	AGKD01109129	PREDICTED: *Esox lucius* testis expressed 10 (tex10), transcript variant X1
1	ESTV_15617_513	Ssa06	–	–	AGKD01044472	PREDICTED: *Neolamprologus brichardi* serine/threonine-protein kinase/endoribonuclease IRE1-like
1	ESTNV_12926_381	Ssa06	29.4	3.4	AGKD01008089	*Salmo salar* nuclear factor interleukin-3-regulated protein
1	ESTV_19580_277	Ssa06	–	–	AGKD01036436	*Salmo salar* THO complex subunit 4 putative
7	ESTNV_14967_182	Ssa07	55	57.9	AGKD01030154	PREDICTED: *Esox lucius* islet amyloid polypeptide (iapp)
1	BASS121_B7_G07_685	Ssa08	22.6	0.7	AGKD01059589	No hits
2	GCR_cBin3425_Ctg1_442	Ssa08	13.3	0.5	AGKD01119363	*Salmo salar* Gastrula zinc finger protein
1	ESTNV_23463_496	Ssa09	–	–	AGKD01043972	*Oncorhynchus mykiss* Glucosamine-6-phosphate isomerase (gnpi)
4	GCR_cBin31103_Ctg1_204	Ssa09	87.4	5.6	AGKD01014092	No hits
1	GCR_cBin28815_Ctg1_209	Ssa10	82.5	8.6	AGKD01091269	No hits
1	ESTV_12965_608	Ssa11	–	–	AGKD01340718	PREDICTED: *Esox lucius* sema domain, transmembrane domain (sema6d)
1	ESTNV_14050_192	Ssa11	61.7	2.1	AGKD01111203	PREDICTED: Esox lucius crystallin, beta B1 (crybb1)
1	ESTNV_27648_141	Ssa11	–	–	AGKD01150671	No hits
1	GCR_cBin47868_Ctg1_188	Ssa11	0.1	0	AGKD01001653	No hits
5	ESTNV_30600_327	Ssa11	12.9	0.2	AGKD01020938	Salmo salar Probable RNA-directed DNA polymerase from transposon BS putative[a]
1	ESTNV_36261_377	Ssa13	71.6	3.8	AGKD01241736	Salmo salar Membrane-associated progesterone receptor component 1 (pgrc1)
1	GCR_cBin5494_Ctg1_253	Ssa16	57.2	4.7	AGKD01163968	No hits
1	ESTV_19974_640	Ssa16–17	–	–	AGKD01011753	Salmo salar mammary gland protein-like (MMADHC)
1	ESTV_17052_125	Ssa16–17	–	–	AGKD01082053	Salmo salar Small ubiquitin-related modifier 3 precursor
1	GCR_cBin49985_Ctg1_148	Ssa19	58.7	9	AGKD01007201	PREDICTED: Esox lucius probable polyketide synthase 1[a]
1	GCR_cBin21476_Ctg1_139	Ssa21	42.2	1.2	AGKD01132253	Salmo salar family with sequence similarity 3, member A (fam3a)[a]
1	GCR_cBin27948_Ctg1_211	Ssa22	7.8	0	AGKD01005678	No hits
6	GCR_cBin108_Ctg1_303	Ssa26	62	1.2	AGKD01005178	No hits
3	ESTNV_28566_397	Ssa29	45.8	0.2	AGKD01044105	Salmo salar 26S protease regulatory subunit S10B (prs10)

[a] Indicates that the WGS contig sequence was used for BLASTn annotation instead of the marker sequence. Female and Male maps in cM, based on Atlantic salmon linkage map described by Lien et al. (2011). Rank based on BayeScan significance levels.

The power to detect selection at a particular locus depends primarily on the number of generations since domestication, the strength of selection and the number of populations included in the study (Karlsson and Moen, 2010; Mäkinen et al., 2015). It has been suggested that if the genetic background of a trait consists of many loci with small additive effects, the selection coefficient for each locus is very small, and so a large number of generations would be required to accumulate a footprint of selection (Pritchard et al., 2010). Atlantic salmon selective breeding has been practiced for the last 45 years, approximately 12 generations, which could be considered as an early stage of domestication. Nevertheless, studies in turbot (*Scophthalmus maximus*) suggest that domestication can generate evidence of selection for some traits after only four generations (Vilas et al., 2015). Certainly, farmed populations of Atlantic salmon have experienced a shorter selection period compared to other livestock species. Even so, differential expression profiles between farmed and wild populations of Atlantic salmon have been reported (Bicskei et al., 2014). The intensity of the selection and the nature of the selected trait must also be taken into consideration; Norwegian domesticated strains have gone through strong selection for growth and against grilsing (i.e., early age of sexual maturation) since they were established (Gjedrem and Baranski, 2010; Gjedrem, 2012). Since the Cermaq broodstock was established in 1995, it too has been selected for these traits. Although selection for growth rate has been carried out intensively, strong, related association signatures have not been detected (Gutierrez et al., 2015). The identification of highly significant outliers within the Cermaq population could indicate that some traits controlled by fewer loci (e.g., late sexual maturation) have been selected and therefore their selection footprints are easier to detect. Breeding programs based on efficient genetic evaluation methods (e.g., the best linear unbiased predictor, BLUP) seem to have had a large effect on the allele frequency in livestock populations (Fontanesi et al., 2015). However, at least some of these differences could also be explained by random events like the emergence a beneficial mutation or genetic drift within the isolated population.

4.3. Genomic regions under selection

The identification of markers as outliers suggest that these could be related to the selection process and therefore, nearby genes could encode proteins associated with the selected traits. Most traits are polygenic, which means that they are controlled by several genes and/or alleles that are each making a contribution to the phenotype. Hence, selection acting on polygenic traits may lead to subtle shifts in allele frequency at many loci. The results presented here and current evidence clearly suggest that numerous regions within the Atlantic salmon genome have experienced this phenomenon (Karlsson et al., 2011; Mäkinen et al., 2015).

The 44 markers that were detected as significant outliers by the three approaches taken (see above) were distributed along 22 of the 29 Atlantic salmon chromosomes. Some of the markers had not been mapped on the currently available linkage map (Lien et al., 2011). However, we were able to assign them to a chromosome based on their sequence and posterior BLAST analysis against the Atlantic salmon genome sequence. Positions of the markers within the chromosomes span regions ranging from 2 to 90 cM in the female map. There are not many markers over-represented in a specific region of the chromosome that would suggest a strong selection. The highest numbers of markers are located on Ssa02, Ssa06 and Ssa11, as shown in Table 4.

Please cite this article as: Gutierrez, A.P., et al., Evidence of recent signatures of selection during domestication in an Atlantic salmon population, Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.12.007

Previous studies analysing signatures of selection in Atlantic salmon have described the presence of outliers on numerous chromosomes (Mäkinen et al., 2015; Vasemägi et al., 2012), but even considering the agreement in the chromosome number, the positions of the markers are not similar. The results obtained from previous studies cannot easily be compared without acknowledging the origin of the populations in question. As previously described, the Cermaq population has a Norwegian origin, and populations analysed in previous studies come from different geographical locations. Moreover, it must be stressed that there are probably differences in nature and the artificial selection processes in farmed populations, which may be reflected in the outcome of the analysis. That being said, this analysis is very conservative when compared to previous findings. Strictly, in terms of QTL analyses, only the studies by Gutierrez et al. (2012, 2014) and Baranski et al. (2010) were carried out on Norwegian populations. Similarities are difficult to compare due to the large number of QTL found for most traits. For example, QTL for growth (or body weight) have been found on most Atlantic salmon chromosomes. Therefore, it would be premature to link growth related QTL to the regions showing the presence of outliers (suggestive selection). On the other hand, it was interesting to note that the only outlier marker on Ssa26 was located on the same Atlantic salmon genome contig as the IPN resistance gene (Houston et al., 2008; Moen et al., 2015).

GO annotation of the most significant markers showed that most genes were linked to biological pathways associated with cellular process, metabolic process and biological regulation. However, no significant over-representation of any particular pathway was observed. Moreover, sequence annotations based on the marker associated sequences or their surrounding genomic regions showed not only many genes with multiple molecular functions, but also many markers that could not be associated to any known genes, at least in the 10 kb surrounding region. However, there were a few markers that seem to be related to a molecular function of importance for selection. For example, many of these have been related to immune or inflammatory response from studies in humans. ESTNV_24005_495 shows sequence similarity with Nod1 protein, implicated in inflammatory response (Fritz et al., 2006), ESTNV_12926_381 is associated with Nfil3, an interleukin with a role in the immune response (Kamizono et al., 2009), ESTV_20616_997 is associated with PU.1, a transcription factor linked to macrophage proliferation (Celada et al., 1996), ESTV_12965_608 is associated with Sema6d a gene with a role in immune response and development (Vadasz and Toubi, 2014) and ESTV_16140_475 is associated with prothymosin α that has an immunoregulatory role during infection (Shiau et al., 2005). In addition, there were other markers that showed an association with genes involved in the regulation of food intake. ESTNV_28438_568 is associated with Sirt5, a gene playing a pivotal role in ammonia detoxification during fasting (Nakagawa et al., 2009) and ESTNV_14967_182 is associated with Amylin, that is believed to play a role in controlling gastric emptying, glucose homeostasis and in the suppression of glucagon release (Cao et al., 2013). ESTNV_35725_1965, is associated with Tex10, a gene with a role in sexual development (Eid et al., 2015) and ESTV_19974_640 which is associated with Mmadhc, a gene involved in vitamin B12 metabolism (Plesa et al., 2011).

We are particularly interested in three markers that were identified as outliers in this analysis (GCR_cBin6804_Ctg1_99 on Ssa02, ESTNV_30600_327 on Ssa11, and ESTNV_36261_377 on Ssa13; Table 4) as they had previously been associated with grilsing in a GWAS study on the Cermaq broodstock (Gutierrez et al., 2015). The GWAS study also indicated that there was a region of Ssa25 associated with grilsing, and recent reports from Norwegian groups (Ayllon et al., 2015; Barson et al., 2015) have shown that variation in *vgll3*, which is located in this region on Ssa25, can account for 33–40% of the grilsing phenotypic. We speculate that there are genes on Ssa02, Ssa11 and Ssa13 that may contribute to the remaining 60–67% of the phenotypic variation. Although the 6.5K SNP array does not have the resolving power to narrow down the genomic region to a specific gene, it is worth noting that one of these markers, ESTNV_36261_377 on Ssa11, is closely linked to *pgrc*1, a gene encoding membrane-associated progesterone receptor component 1, that is implicated in hormone metabolism (Rohe et al., 2009). We suggest that *pgrc*1 is a candidate gene for determining grilsing in Atlantic salmon.

5. Conclusions

We found significant evidence for signatures of selection in the Cermaq farmed population of Atlantic salmon. Even though the number of markers identified as outliers is low (~1%), some of the markers reach high levels of significance and are associated with molecular functions that could be related to the selection process of particular traits. There was little similarity in our observations compared to previous studies arguing for little evidence of selection of Atlantic salmon populations due to the short selection period (~12 generations). Further studies are needed to determine the nature of these signatures and the relevance of these loci in the adaptation process associated to selection in the Cermaq population. This study is the first to find strong signatures of adaptation to farming environments and strong selection pressures in the genome of an Atlantic salmon farmed population.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2015.12.007.

Author contributions

APG performed the analyses. APG, JMY & WSD conceived the study, analysed the results and wrote the article.

Acknowledgements

We would like to thank to the authors of the Bourret et al. (2013) article for making their genotype data publicly available. This study was funded through an NSERC strategic grant (STPGP/381479-2009) to WSD.

References

Ai, H., Huang, L., Ren, J., 2013. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. PLoS One 8, e56001.
Ashburner, M., Ball, C.A., Blake, J.A., et al., 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.
Ayllon, F., Kjaerner-Semb, E., Furmanek, T., et al., 2015. The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar* L.) males. PLoS Genet. http://dx.doi.org/10.1371/journal.pgen.1005628.
Baranski, M., Moen, T., Vage, D., 2010. Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). Genet. Sel. Evol. 42, 17.
Barson, N.J., Aykanat, T., Hindar, K., et al., 2015. Sex-dependence dominance at a single locus maintains variation in age at maturity in salmon. Nature http://dx.doi.org/10.1038/nature16062.
Beaumont, M.A., Nichols, R.A., 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. Biol. Sci. 263, 1619–1626.
Bicskei, B., Bron, J., Glover, K., Taggart, J., 2014. A comparison of gene transcription profiles of domesticated and wild Atlantic salmon (*Salmo salar* L.) at early life stages, reared under controlled conditions. BMC Genomics 15, 884.
Bijma, P., 2012. Long-term genomic improvement — new challenges for population genetics. J. Anim. Breed. Genet. 129, 1–2.
Bourret, V., Kent, M.P., Primmer, C.R., et al., 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Mol. Ecol. 22, 532–551.
Brenna-Hansen, S., Li, J., Kent, M.P., et al., 2012. Chromosomal differences between European and North American Atlantic salmon discovered by linkage mapping and supported by fluorescence in situ hybridization analysis. BMC Genomics 13, 432.
Cao, P., Abedini, A., Raleigh, D.P., 2013. Aggregation of islet amyloid polypeptide: from physical chemistry to cell biology. Curr. Opin. Struct. Biol. 23, 82–89.
Celada, A., Borràs, F.E., Soler, C., et al., 1996. The transcription factor PU.1 is involved in macrophage proliferation. J. Exp. Med. 184, 61–69.
Conesa, A., Götz, S., García-Gómez, J.M., et al., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.
Davidson, W.S., Koop, B.F., Jones, S., et al., 2010. Sequencing the genome of the Atlantic salmon (*Salmo salar*). Genome Biol. 11, 403.

Eid, W., Opitz, L., Biason-Lauber, A., 2015. Genome-wide identification of CBX2 targets: insights in the human sex development network. Mol. Endocrinol. 29, 247–257.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567.

Excoffier, L., Hofer, T., Foll, M., 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103, 285–298.

Falconer, D.S., Mackay, T.F., 1996. Introduction to Quantitative Genetics. Longman, Harlow. UK.

Falush, D., Stephens, M., Pritchard, J.K., 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7, 574–578.

Foll, M., Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180, 977–993.

Fontanesi, L., Schiavo, G., Scotti, E., et al., 2015. A retrospective analysis of allele frequency changes of major genes during 20 years of selection in the Italian Large White pig breed. J. Anim. Breed. Genet. 132, 239–246.

Fourcade, Y., Chaput-Bardy, A., Secondi, J., Fleurant, C., Lemaire, C., 2013. Mol. Ecol. 22, 2065–2073.

Fritz, J.H., Ferrero, R.L., Philpott, D.J., Girardin, S.E., 2006. Nod-like proteins in immunity, inflammation and disease. Nat. Immunol. 7, 1250–1257.

Gjedrem, T., 2012. Genetic improvement for the development of efficient global aquaculture: a personal opinion review. Aquaculture 344–349, 12–22.

Gjedrem, T., Baranski, M., 2010. Selective Breeding in Aquaculture: An Introduction. Springer.

Glover, K.A., Otterå, H., Olsen, R.E., et al., 2009. A comparison of farmed, wild and hybrid Atlantic salmon (*Salmo salar* L.) reared under farming conditions. Aquaculture 286, 203–210.

Guo, S.W., Thompson, E.A., 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48, 361–372.

Gutierrez, A.P., Lubieniecki, K.P., Davidson, E.A., et al., 2012. Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5K SNP array. Aquaculture 358–359, 61–70.

Gutierrez, A.P., Lubieniecki, K.P., Fukui, S., et al., 2014. Detection of quantitative trait loci (QTL) related to grilsing and late sexual maturation in Atlantic salmon (*Salmo salar*). Mar. Biotechnol. (N.Y.) **16**, 103–110.

Gutierrez, A.P., Yáñez, J.M., Fukui, S., Swift, B., Davidson, W.S., 2015. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). PLoS One 10, e0119730.

Houston, R.D., Haley, C.S., Hamilton, A., et al., 2008. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). Genetics 178, 1109–1115.

Jombart, T., 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.

Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11, 94.

Kamizono, S., Duncan, G.S., Seidel, M.G., et al., 2009. Nfil3/E4bp4 is required for the development and maturation of NK cells in vivo. J. Exp. Med. 206, 2977–2986.

Karlsson, S., Moen, T., 2010. The power to detect artificial selection acting on single loci in recently domesticated species. BMC Res. Notes 3, 232.

Karlsson, S., Moen, T., Lien, S., Glover, K.A., Hindar, K., 2011. Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Mol. Ecol. Resour. 11, 247–253.

Kauer, M.O., Dieringer, D., Schlötterer, C., 2003. A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of *Drosophila melanogaster*. Genetics 165, 1137–1148.

Kent, M.P., Hayes, B., Xiang, Q., et al., 2009. Development of 16.5K SNP-Chip for Atlantic salmon. Plant & Animal Genomes XVII Conference, San Diego, CA.

Kijas, J.W., Lenstra, J.A., Hayes, B., et al., 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10, e1001258.

Kraemer, P., Gerlach, G., 2013. R Package "Demerelate" (cran.r-project.org.) pp. 1–33.

Lien, S., Gidskehaug, L., Moen, T., et al., 2011. A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. BMC Genomics 12, 615.

Lopez, M.E., Neira, R., Yáñez, J.M., 2015. Applications in the search for genomic selection signatures in fish. Front. Genet. 5.

Lubieniecki, K., Jones, S., Davidson, E., et al., 2010. Comparative genomic analysis of Atlantic salmon, *Salmo salar*, from Europe and North America. BMC Genet. 11, 105.

Mäkinen, H., Vasemägi, A., McGinnity, P., Cross, T.F., Primmer, C.R., 2015. Population genomic analyses of early-phase Atlantic Salmon (*Salmo salar*) domestication/captive breeding. Evol. Appl. 8, 93–107.

Martinez, V., Dettleff, P., Lopez, P., et al., 2013. Assessing footprints of selection in commercial Atlantic salmon populations using microsatellite data. Anim. Genet. 44, 223–226.

Moen, T., Torgersen, J., Santi, N., et al., 2015. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. Genetics 200, 1313–1326.

Nakagawa, T., Lomb, D.J., Haigis, M.C., Guarente, L., 2009. SIRT5 Deacetylates carbamoyl phosphate synthetase 1 and regulates the urea cycle. Cell 137, 560–570.

Narum, S.R., Hess, J.E., 2011. Comparison of FST outlier tests for SNP loci under selection. Mol. Ecol. Resour. 11, 184–https://www.google.com.ph/webhp?hl=en194.

Nei, M., 1972. Genetic distances between populations. Am. Nat. 106, 283–292.

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., Clark, A.G., 2007. Recent and ongoing selection in the human genome. Nat. Rev. Genet. 8, 857–868.

Nosil, P., Funk, D.J., Ortiz-Barrientos, D., 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18, 375–402.

Pennings, P.S., Hermisson, J., 2006. Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet. 2, e186.

Perrier, C., Bourret, V., Kent, M.P., Bernatchez, L., 2013. Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. Mol. Ecol. 22, 5577–5593.

Phillips, R.B., Keatley, K.A., Morasch, M.R., et al., 2009. Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). BMC Genet. 10, 46.

Plesa, M., Kim, J., Paquette, S.G., et al., 2011. Interaction between MMACHC and MMADHC, two human proteins participating in intracellular vitamin B12 metabolism. Mol. Genet. Metab. 102, 139–148.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Pritchard, J.K., Pickrell, J.K., Coop, G., 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 20, R208–R215.

Qanbari, S., Strom, T.M., Haberer, G., et al., 2012. A high resolution genome-wide scan for significant selective sweeps: an application to pooled sequence data in laying chickens. PLoS One 7, e49525.

Qanbari, S., Pausch, H., Jansen, S., et al., 2014. Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet. 10, e1004148.

Ramey, H., Decker, J., McKay, S., et al., 2013. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genomics 14, 382.

Rohe, H.J., Ahmed, I.S., Twist, K.E., Craven, R.J., 2009. PGRMC1 (progesterone receptor membrane component 1): a targetable protein with multiple functions in steroid signaling, P450 activation and drug binding. Pharmacol. Ther. 121, 14–19.

Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.

Shiau, A.-L., Chen, C.-C., Yo, Y.-T., et al., 2005. Enhancement of humoral and cellular immune responses by an oral *Salmonella choleraesuis* vaccine expressing porcine prothymosin α. Vaccine 23, 5563–5571.

Smith, J.M., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23, 23–35.

Teshima, K.M., Coop, G., Przeworski, M., 2006. How reliable are empirical genomic scans for selective sweeps? Genome Res. 16, 702–712.

Vadasz, Z., Toubi, E., 2014. Semaphorins: their dual role in regulating immune-mediated diseases. Clin. Rev. Allergy Immunol. 47, 17–25.

Vasemägi, A., Nilsson, J., Primmer, C.R., 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol. Biol. Evol. 22, 1067–1076.

Vasemägi, A., Nilsson, J., McGinnity, P., et al., 2012. Screen for footprints of selection during domestication/captive breeding of Atlantic salmon. Comp. Funct. Genomics 2012, 14.

Verspoor, E., Stradmeyer, L., Nielsen, J.L., 2007. The Atlantic Salmon: Genetics, Conservation, and Management. Blackwell Publishing Ltd., Oxford.

Vilas, R., et al., 2015. A genome scan for candidate genes involved in the adaptation of turbot (*Scophthalmus maximus*). Mar. Genomics 23, 77–86.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Zueva, K.J., Lumme, J., Veselov, A.E., et al., 2014. Footprints of directional selection in wild Atlantic salmon populations: evidence for parasite-driven evolution? PLoS One 9, e91672.

Please cite this article as: Gutierrez, A.P., et al., Evidence of recent signatures of selection during domestication in an Atlantic salmon population, Mar. Genomics (2015), http://dx.doi.org/10.1016/j.margen.2015.12.007